ARTICLE IN PRESS

Journal of Complexity ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of Complexity

journal homepage: www.elsevier.com/locate/jco

The exact information-based complexity of smooth convex minimization ☆

Yoel Drori

Google Inc., 1600 Amphitheatre Parkway, Mountain View, CA 94043, United States

ARTICLE INFO

Article history:
Received 11 June 2016
Accepted 9 November 2016
Available online xxxx

Keywords:
Convex optimization
Complexity
Rate of convergence
Information-based complexity

ABSTRACT

We obtain a new lower bound on the information-based complexity of first-order minimization of smooth and convex functions. We show that the bound matches the worst-case performance of the recently introduced Optimized Gradient Method (Drori and Teboulle, 2013; Kim and Fessler, 2015), thereby establishing that the bound is tight and can be realized by an efficient algorithm. The proof is based on a novel construction technique of smooth and convex functions.

1. Introduction

The problem of smooth and convex minimization plays a key role in a various range of applications, including signal and image processing, communications, machine learning, and many more. Some of the most successful approaches for solving these problems are first-order methods, i.e., algorithms that are only allowed to gain information on the objective by evaluating its value and gradient. The benefit of limiting the amount of accessible information is that these algorithms generally involve very cheap and simple computational iterations, making them suitable for tackling large scale problems. This benefit, however, comes with a price: first-order methods, in general, require considerable computational effort in order to reach highly accurate solutions, making them practical when only moderate accuracy is sufficient.

As the scale of modern problems grows and finding efficient algorithms becomes increasingly important, a natural question that arises, and will be the main focus of this paper, is finding the level of

☆ Communicated by E. Novak.

E-mail address: dyoel@google.com.

http://dx.doi.org/10.1016/j.jco.2016.11.001
0885-064X/

Please cite this article in press as: Y. Drori, The exact information-based complexity of smooth convex minimization, Journal of Complexity (2016), http://dx.doi.org/10.1016/j.jco.2016.11.001

accuracy that can be attained by first-order methods using a bounded computational effort. Note that there is some difficulty in answering this question that originates from the fact that the computational effort of a first-order method consists of two parts: the effort in choosing the points where the objective is to be evaluated (called the *search points*) and the effort in calculating the objective value and gradient at these points. Observing that the evaluation of the objective and its gradient often dominates the computational effort of the computation and following the theory of information-based complexity introduced in [8], we resolve this issue by measuring the computational effort of an algorithm by the number of times it evaluated the objective and its gradient, neglecting the effort required for choosing the search points.

To put these concepts in more precise terms, consider the following unconstrained problem

$$(P)\, f^* = \min_{x\in\mathbb{R}^d} f(x),$$

where f is a smooth and convex function. A *first-order* optimization method is an iterative algorithm that approximates the solution of (P), where it is only allowed to gain information on the objective f via a first-order oracle, $\mathcal{O}_f$, that is, a subroutine which given a point in $\mathbb{R}^d$, returns the value of the objective and its gradient at that point. In addition, since the problem of unconstrained minimization is invariant under translations, we also assume that the algorithm is provided with a reference (or starting) point $x_0 \in \operatorname{dom}(f)$ that is often assumed to be "not too far" from an optimal solution. We call the pair $(\mathcal{O}_f, x_0)$ a *problem instance*, and for a first-order method A we denote the approximate solution generated by algorithm when applied on this problem instance by $A(\mathcal{O}_f, x_0)$.[1]

Within the setting considered above, a commonly used criterion for measuring the accuracy of an approximate solution is the *absolute inaccuracy* criterion, which quantifies the accuracy of an approximate solution ξ for a problem instance $(\mathcal{O}_f, x_0)$ by the value of $f(\xi)-f^*$ (for alternative criteria see e.g., [7, Section 3.3]). Under this criterion, the *efficiency estimate* of a first-order method A over some given set of problem instances $\mathcal{I}$ is defined as the worst-case absolute inaccuracy of A, i.e.,

$$\varepsilon(A;\mathcal{I}) := \sup_{(\mathcal{O}_f,x_0)\in\mathcal{I}} f(A(\mathcal{O}_f,x_0)) - f^*.$$

We can now put the main concept addressed in this paper in formal terms: denoting by $\mathcal{A}_N$ the set of all first-order methods that perform at most $N \in \mathbb{N}$ calls to their first-order oracle, the *minimax risk function* [5] associated with $\mathcal{I}$ is defined as the infimal efficiency estimate that a first-order method can attain over $\mathcal{I}$ as a function of the computational effort N:

$$\mathcal{R}_{\mathcal{I}}(N) := \inf_{A\in\mathcal{A}_N} \varepsilon(A;\mathcal{I}).$$

Note that the classical notion of *information-based complexity* of the set $\mathcal{I}$ can be identified as the inverse to the risk function,

$$\mathcal{C}_{\mathcal{I}}(\varepsilon) := \min\{N : \mathcal{R}_{\mathcal{I}}(N) \le \varepsilon\},$$

i.e., the minimal computational effort needed by a first-order method in order to reach a given worst-case accuracy level, however, in the following we express our results using the minimax risk function as it proves to be more convenient.

Clearly, in order to establish an upper bound on the minimax risk of a class it is sufficient to find an upper bound on the efficiency estimate of a single first-order method (the main problem here being the identification of a good algorithm). On the other hand, establishing lower bounds on the minimax risk requires a more involved analysis, as the bound needs to hold for any first-order method. Several approaches appear in the literature for establishing lower-bounds, including resisting oracles [9], construction of a "worst-case" function [5,10], and reduction to statistical problems [1,11,13].

Note that existing works on information-based complexity focus mainly on establishing order of magnitude bounds, where less attention is paid to absolute constants. Nevertheless, the exact minimax risk was established for several important classes of problems, some of which are detailed below.

[1] In order to simplify the presentation, we assume A has a built-in stopping criterion that does not depend on external input.

Please cite this article in press as: Y. Drori, The exact information-based complexity of smooth convex minimization, Journal of Complexity (2016), http://dx.doi.org/10.1016/j.jco.2016.11.001